

Sparse-TDA: Sparse Realization of Topological Data Analysis for Multi-Way Classification

Wei Guo, *Student Member, IEEE*, Krithika Manohar, Steven L. Brunton and Ashis G. Banerjee, *Member, IEEE*

Abstract—Topological data analysis (TDA) has emerged as one of the most promising techniques to reconstruct the unknown shapes of high-dimensional spaces from observed data samples. TDA, thus, yields key shape descriptors in the form of persistent topological features that can be used for any supervised or unsupervised learning task, including multi-way classification. Sparse sampling, on the other hand, provides a highly efficient technique to reconstruct signals in the spatial-temporal domain from just a few carefully-chosen samples. Here, we present a new method, referred to as the Sparse-TDA algorithm, that combines favorable aspects of the two techniques. This combination is realized by selecting an optimal set of sparse pixel samples from the persistent features generated by a vector-based TDA algorithm. These sparse samples are selected using a low-rank matrix representation of persistent features. We show that the Sparse-TDA method demonstrates promising performance on three benchmark problems related to human posture recognition and image texture classification.

Index Terms—Topological data analysis, sparse sampling, multi-way classification, human posture data, image texture data.

I. INTRODUCTION

MULTI-way or multi-class classification, where the goal is to correctly predict one out of K classes for any data sample, poses one of the most challenging problems in supervised machine learning. However, a large number of real-world sensing problems in a variety of domains such as computer vision, robotics and remote diagnostics, do consist of multiple classes. Examples include human face recognition for surveillance, object detection for mobile robot navigation, and critical equipment condition monitoring for preventive maintenance. The number of classes in these problems often exceeds ten and sometimes goes up to a hundred depending on the complexity of the sensed system or environment and the number and types of sensor modalities.

While a whole host of techniques such as artificial neural networks, decision trees, naïve Bayes, nearest neighbors, and support vector machines (SVMs) have been successfully applied for binary classification problems, extensions of these techniques have had mixed success in addressing multi-way classification problems with more than a few classes. Other approaches involving hierarchical classification or transformation

to binary classification have not been particularly successful either. The success rates diminish further in the absence of a large number of data samples for each of the labeled classes. The primary reason is that all of these methods encounter difficulties in selecting the right set of distinguishing features among the different classes.

Recent research has started investigating completely new techniques for multi-way classification that attempt to better understand the structure of the underlying high-dimensional sample space. One such class of techniques is topological data analysis, or TDA in short. TDA represents the unknown sample space in the form of persistent shape descriptors that are coordinate free and deformation invariant. Thus, the descriptors define topological features and yield insights regarding suitable feature selection.

Another critical tool facilitating multi-way classification is the feature-driven sparse sampling of high-dimensional data. Observations are typically sparse in a transform basis of informative features, so measurements can be optimally chosen to enhance discriminating features in the data. This permits heavily subsampled inputs for downstream classifiers, which drastically reduces the burdens of sample acquisition, processing and storage without sacrificing performance.

Here, we bring together the two research areas of TDA and sparse sampling in the context of multi-way classification. In particular, we leverage sparse sampling for optimal feature selection once the features are extracted using a state-of-the-art TDA method in challenging computer vision problems. The problems comprise three benchmark data sets pertaining to 3D meshes of synthetic and real human postures and textured images, respectively. We call our new method the Sparse-TDA algorithm and show that it achieves comparable accuracy as the TDA method with significantly lower training times. Thus, our method opens up a new direction in making online multi-way classification practically feasible.

II. RELATED WORK

Over the past decade or so, an increasing interest in utilizing tools from algebraic topology to extract insights from high dimensional data has given rise to the field of TDA. The successful applications of TDA have spanned a large number of areas, ranging from computer vision [1] to medical imaging [2], biochemistry [3], neuroscience [4] and materials science [5]. A predominant tool in TDA is persistent homology, which tracks the evolution of the topological features in a multi-scale manner to avoid information loss [6], [7]. The multi-scale information is summarized by the persistence diagram

W. Guo is with the Department of Industrial & Systems Engineering, University of Washington, Seattle, WA, 98195 USA e-mail: weig@uw.edu.

K. Manohar is with the Department of Applied Mathematics, University of Washington, Seattle, WA, 98195 USA e-mail: kmanohar@uw.edu.

S. L. Brunton is with the Department of Mechanical Engineering, University of Washington, Seattle, WA, 98195 USA e-mail: sbrunton@uw.edu.

A. G. Banerjee is with the Department of Industrial & Systems Engineering and Department of Mechanical Engineering, University of Washington, Seattle, WA, 98195 USA e-mail: ashishb@uw.edu.

(PD), a multiset of points in \mathbb{R}^2 that encodes the lifetime (i.e., persistence) of the features.

More recently, researchers have started utilizing TDA for machine learning problems. Pachauri et al. [8] first computed a Gaussian kernel to estimate the density of points on a regular grid for each rasterized PD, and fed the discrete density estimation as a vector into an SVM classifier without any feature selection. However, their method did not establish the stability of the kernel-induced vector representation. Reininghaus et al. [1] then designed a stable multi-scale kernel for PDs motivated by scale-space theory as will be described in the next Section. Experiments on three benchmark data sets showed that this method greatly outperformed an alternative approach based on persistence landscape [9], a popular statistical treatment of TDA. Similar to this work, Kusano et al. [10] proposed a stable persistence weighted Gaussian kernel, allowing one to control the effect of persistence. However, the computational complexity of both the kernel-based methods for calculating the Gram matrix is $O(m^2n^2)$ if there are n PDs for training and the PDs contain at most m points, which can be quite expensive for many practical applications.

To enable large-scale computations with PDs, recent methods have mapped each PD to a stable vector to allow direct use of vector-based learning methods. For example, Adams et al. [11] constructed vectors by discretizing the weighted sum of probability distributions centered at each point in transformed PDs. Carrière et al. [12] rearranged the entries of the distance matrix between points in a PD and Bonis et al. [13] adopted a pooling scheme to construct the vectors.

Existing sparse sampling techniques leverage empirical interpolation methods (EIM) [14]–[16] to identify discriminating measurements using low-rank spanning features, such as obtained by principal components analysis (PCA). This strategy has been employed for sparse classification of dynamical regimes within libraries of learned PCA features [17], [18]. In image classification using linear discriminant analysis (LDA), optimized sparse pixel locations were determined that map into the discriminating subspace in PCA coordinates [19].

In this work, we employ the vector representation from [11] and integrate with a sparse sampling method using QR pivots to identify discriminative features in the presence of noisy and redundant information to further improve classifier training time and sometimes prediction accuracy.

III. KERNEL-BASED TDA METHOD FOR MULTI-WAY CLASSIFICATION

We first introduce the basic terminology of any TDA method by formally defining the mathematical operations and stating the key results. We then use these definitions and results to summarize the multi-scale kernel TDA method. While the kernel TDA method is directly adopted from [1], we outline it here for the sake of completeness.

A. Terminology

Filtration: In computational topology, the “shapes” of data (e.g., point clouds, shapes, images) are frequently described by simplicial (or cubical) complexes. A filtration of a finite

simplicial (or cubical) complex S is a sequence of simplicial complexes S_1, \dots, S_r such that $S_1 \subseteq \dots \subseteq S_r = S$. A common way to generate a filtration is to consider the sublevel sets $S_t = f^{-1}([-\infty, t])$ of a descriptor function $f : \mathbb{X} \rightarrow \mathbb{R}$ on a topological space \mathbb{X} indexed by a parameter $t \in \mathbb{R}$.

Persistent homology: As a prevalent tool in TDA, persistent homology is an algebraic approach that quantifies topological features during a filtration of “shapes” [6], [7]. Accordingly, given a topological space \mathbb{X} and a descriptor function $f : \mathbb{X} \rightarrow \mathbb{R}$, persistent homology essentially studies the topological changes of the sublevel sets $\mathbb{X}_t = f^{-1}([-\infty, t])$ as t increases from $-\infty$ to ∞ . During filtration, topological features appear and disappear at different scales that are referred to as the *birth* and *death* times of the features. The short-lived features are considered as noise terms.

Persistence diagram (PD): A PD is a concise summary of the topological information captured by persistent homology. From a geometric perspective, the topological features are interpreted as l -dimensional holes, e.g., connected components as 0-dimension holes, tunnels as 1-dimensional holes and voids as 2-dimensional holes. Thus, a l -dimensional PD is a collection of points in \mathbb{R}^2 , where each point (a, b) represents a l -dimensional hole that is born at time a and filled at time b .

Stability: A critical property of PDs is their stability with respect to input noise [20], [21]. A general metric associated with PDs is the p -Wasserstein distance. The p -Wasserstein distance between two PDs F and G is defined by

$$d_{W,p}(F, G) = \left(\inf_{\gamma} \sum_{x \in F} \|x - \gamma(x)\|_p^p \right)^{\frac{1}{p}}, \quad 1 \leq p \leq \infty, \quad (1)$$

where $\gamma : F \rightarrow G$ is over all bijections from the points in F to the points in G .

Let X be a compact triangulable metric space and $f, g : X \rightarrow \mathbb{R}$ be two tame Lipschitz functions with the corresponding PDs $D_l(f)$ and $D_l(g)$ for each dimension l . It has been prove that assuming X satisfies a weak condition (see details in [21]), there exist constants $q \geq 1$ and C , which depend on X and the Lipschitz constants of f and g , such that

$$d_{W,p}(D_l(f), D_l(g)) \leq C \|f - g\|_{\infty}^{1-\frac{q}{p}}, \quad p \geq q. \quad (2)$$

This upper bound on $d_{W,p}(D_l(f), D_l(g))$ implies that a PD $D_l(f)$ is stable with respect to the p -Wasserstein distance under small perturbations of f .

Kernel: Given a set \mathcal{X} , $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is a kernel if there exists a Hilbert space \mathcal{H} , called *feature space*, and a map $\Phi : \mathcal{X} \rightarrow \mathcal{H}$ such that

$$k(x, x') = \langle \Phi(x), \Phi(x') \rangle_{\mathcal{H}} \quad (3)$$

for all $x, x' \in \mathcal{X}$. In machine learning, a kernel represents a similarity measure between the samples, and Φ is called its *feature map*. A kernel satisfying Eq. (3) is also symmetric and positive definite [22].

B. Multi-scale Kernel TDA

In [1], Reininghaus et al. devise the *persistence scale space* kernel on the set of PD \mathcal{D} as a multi-scale kernel via a feature

map $\Phi_\sigma : \mathcal{D} \rightarrow L_2(\Omega)$, where $\Omega = \{x = (x_1, x_2) \in \mathbb{R}^2 : x_2 \geq x_1\}$ denotes the space above the diagonal. Given a PD $D \in \mathcal{D}$, the feature map Φ_σ is the solution of a heat diffusion problem with a Dirichlet boundary condition on the diagonal:

$$\Phi_\sigma(D) = \frac{1}{4\pi\sigma} \sum_{y \in D} \left(e^{-\frac{\|x-y\|^2}{4\sigma}} - e^{-\frac{\|x-\bar{y}\|^2}{4\sigma}} \right), \quad (4)$$

where $\bar{y} = (b, a)$ is the mirror image of $y = (a, b)$ across the diagonal. The map then yields the kernel $k_\sigma : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$ in a closed form as

$$\begin{aligned} k_\sigma(F, G) &= \langle \Phi_\sigma(F), \Phi_\sigma(G) \rangle_{L_2(\Omega)} \\ &= \frac{1}{8\pi\sigma} \sum_{\substack{y \in F \\ z \in G}} \left(e^{-\frac{\|y-z\|^2}{8\sigma}} - e^{-\frac{\|y-\bar{z}\|^2}{8\sigma}} \right) \end{aligned} \quad (5)$$

for $\sigma > 0$ and $F, G \in \mathcal{D}$, which has been shown to be 1-Wasserstein stable. Further, note that because the summation in Eq. (5) is carried out over all pairwise combinations of the points in the PDs F and G , evaluation of the kernel requires $O(|F||G|)$ time, where $|F|$ and $|G|$ denote the number of points in F and G , respectively.

IV. SPARSE-TDA METHOD

We now introduce a vector representation of a PD, termed a persistence image (PI), presented in [11]. Since our Sparse-TDA method will combine PI-based TDA with sparse sample selection, we first summarize the sparse sampling method before describing the combination.

A. Optimized Sparse Sample Selection

Samples are selected from low-rank feature vectors using the pivoted QR matrix factorization, also known as Q-DEIM [16]. Assuming observations are columns in a data matrix \mathbf{X} , we first obtain a sparsifying basis for its columns from the left singular vectors of its singular value decomposition

$$\mathbf{X} \approx \mathbf{U}_s \mathbf{\Sigma}_s \mathbf{V}_s^T, \quad (6)$$

where s is the rank of the truncated basis and the number of desired measurements. The column pivoted QR factorization of \mathbf{U}_s^T optimizes a row permutation of the basis that is numerically well-conditioned by expressing \mathbf{U}_s^T as a product of unitary and upper-triangular matrices \mathbf{Q} and \mathbf{R}

$$\mathbf{U}_s^T \mathbf{\Pi}^T = \mathbf{Q} \mathbf{R}. \quad (7)$$

The optimized row permutation is $\mathbf{\Pi} \mathbf{U}_s$, and the sparse measurements of an observation \mathbf{x} are given by $\tilde{\mathbf{x}} = \mathbf{\Pi} \mathbf{x}$.

B. Combining Sparse Feature Selection with Persistence Images

Let $\mathcal{D} = \{D_i \mid i = 1, \dots, n\}$ be a training set of PDs. To construct a PI from a given PD D_i [11], D_i is first transformed from birth-death coordinates to birth-persistence coordinates. Let $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be the linear transformation,

$$T(x, y) = (x, y - x). \quad (8)$$

A persistence surface $\rho_{D_i} : \mathbb{R}^2 \rightarrow \mathbb{R}$ on $T(D_i)$ is defined by

$$\rho_{D_i}(z) = \sum_{u \in T(D_i)} f(u) g_u(z) \quad (9)$$

where $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a non-negative weighting function that is zero along the horizontal axis, continuous, and piecewise differentiable; $g_u : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a probability function with mean $u = (u_x, u_y) \in \mathbb{R}^2$ and variance σ^2 .

In our experiments, the linear weighting function is

$$f(u) = \frac{u_y}{u_y^*}, \quad (10)$$

where $u_y^* = \max_{i=1, \dots, n} \max_{u \in D_i} u_y$. The form of the non-linear weighting function is inspired by the weighting function used in [10] and chosen as

$$f(u) = \arctan(cu_y). \quad (11)$$

where $c = (\text{median}_{i=1, \dots, n} \text{median}_{u \in D_i} u_y)^{-1}$. We choose g_u to be the Gaussian distribution, i.e.,

$$g_u(z) = \frac{1}{2\pi\sigma^2} e^{-[(z_x - u_x)^2 + (z_y - u_y)^2]/2\sigma^2}. \quad (12)$$

where $z = (z_x, z_y)$. Then the PI, a matrix of pixel values, is obtained by calculating the integral of ρ_{D_i} on each grid box from discretization,

$$I(\rho_{D_i}) = \iint \rho_{D_i}(z_x, z_y) dz_x dz_y. \quad (13)$$

PI has also been proven to be 1-Wasserstein stable. Let the grid resolution for each PI be $r \times r$. We reshape each PI to be a vector $\mathbf{x}_i \in \mathbb{R}^{r^2}$, $i = 1, \dots, n$. Our sparse sample selection method is unsupervised, where we then stack the vectors from all the classes columnwise into a matrix as

$$\mathbf{X} \in \mathbb{R}^{r^2 \times n} = [\mathbf{x}_1 \mid \mathbf{x}_2 \mid \dots \mid \mathbf{x}_n]. \quad (14)$$

Assume that the number of desired features is s . Applying the sparse sampling method on \mathbf{X} , we obtain the row index of s optimal feature locations and the *reduced feature vectors* $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_n \in \mathbb{R}^s$ for the corresponding classifiers.

V. RESULTS

We now discuss the performance of our Sparse-TDA method on three benchmark computer vision data sets. The data sets are explained first, followed by illustrations of the selected features and quantitative comparisons of our method with the multi-scale kernel TDA method. The illustrations and comparison results show the usefulness of the method on challenging multi-way classification problems.

A. Data Sets

For shape classification, SHREC'14 synthetic and real data sets are used, given in the format of triangulated 3D meshes [23]. The synthetic set contains meshes from five males, five females and five children in 20 different poses, while the real set consists of 20 males and 20 females in 10 different poses.

For texture recognition, we use the Outex_TC_00000 data set [24]. This data set contains 480 images equally categorized into 24 classes and provides 100 predefined 50/50 training/testing splits. During preprocessing, we downsample the original images to 32×32 pixel images as done in the multi-scale kernel TDA method.

B. Feature Selection

We first follow the same procedure performed in the multi-scale kernel TDA method to obtain PDs. For SHREC'14 data sets, we compute the heat kernel signature [25] on the surface mesh of each object and then compute the 1-dimensional PDs using DIPA¹. For the OuTeX data set, we take the sign component of the completed local binary pattern operator [26] as the descriptor function. Then we generate the 0-dimensional PDs from the filtration of its rotation-invariant version with $P = 8$ neighbors and radius $R = 1$.

To generate the PIs, we set the resolution to be 30×30 for all three data sets. In fact, the classification accuracy is fairly robust to the choice of resolution [11]. We also set σ to be 0.15, 0.0001 and 0.02 for SHREC'14 synthetic, SHREC'14 real and OuTeX data sets, respectively. Fig. 1 shows representative PIs for three different classes in all of our benchmark data sets. Noticeable differences are observed among the PIs for each of the three data sets, although the differences are most pronounced for the SHREC'14 synthetic data set, reasonably clear for the SHREC'14 synthetic data set, and less evident for the OuTeX data set. These differences in the pixel values of the PIs form the distinguishing class features from which an optimal set is selected by QR pivots.

TABLE I
SIZE OF THE TRAINING SETS AND ENERGY CONTAINED IN THE 100
OPTIMIZED PIXEL LOCATIONS (IN %)

Data Set	Training Size	Sparse-TDA	
		Linear weighting	Non-linear weighting
SHREC'14 Synthetic	210	100 \pm 0	100 \pm 0
SHREC'14 Real	280	97.8 \pm 0.13	97.6 \pm 0.12
OuTeX Texture	240	89.5 \pm 0.12	91.0 \pm 0.12

Pixels are selected using the QR pivots of the first $s = 100$ left singular vectors in each dataset, therefore Sparse-TDA accuracy is affected by how much energy or variance in each training set is explained within the first 100 basis vectors U_s . The energy of each training set, tabulated in Table I, is the normalized sum of the first 100 singular values (diagonal entries of Σ). For instance, it can be seen that the SHREC'14 synthetic data set is intrinsically lower rank.

C. Classification Performance

We feed the reduced feature vectors for training into a soft margin C-SVM classifier with a radial basis function (RBF) kernel, implemented in LIBSVM [27], for each data set. The cost factor C and kernel parameter γ are tuned based on a grid search using 10-fold cross-validation on the training data. We start a coarse grid search with exponentially growing sequences of C and γ first, thereafter proceeding with finer grid searches in the vicinity of the optimal region yielded by the previous grid search. Each grid search includes a total of 50 pairs of (C, γ) values which are used to apply the training model to the reduced feature vectors of the test

set. Results are reported based on 30 runs for each case with the exception of those presented for the OuTeX data set in Tables I-III, which are based on 100 runs.

Table II compares the classification accuracy of both the variants of our Sparse-TDA method with the multi-scale kernel TDA method. Note that the two SHREC'14 data sets are partitioned into 70/30 training/testing samples, whereas the OuTeX set is partitioned into 50/50 training/testing samples. The number of samples for each class is approximately the same in all the training sets. Overall, the accuracy results are comparable for all the three benchmark data sets. However, consistent with the differences observed in the PIs among the classes, both our linear and non-linear weighting methods perform slightly better than the kernel TDA method for the SHREC'14 real data set. On the other hand, our method is marginally worse than the kernel TDA method for both the SHREC'14 synthetic and OuTeX data sets, even though the accuracy increases slightly using non-linear weighting.

TABLE II
COMPARISON OF CLASSIFICATION ACCURACY (IN %) WITH THE SAME
TRAINING AND TEST DATA SET SPLIT AS IN THE ORIGINAL MULTI-SCALE
KERNEL TDA ARTICLE

Data Set	Sparse-TDA		Kernel TDA
	Linear weighting	Non-linear weighting	
SHREC'14 Synthetic	91.7 \pm 2.6	93.6 \pm 2.7	98.4 \pm 1.7
SHREC'14 Real	67.3 \pm 4.1	67.9 \pm 4.6	65.9 \pm 4.4
OuTeX Texture	62.5 \pm 2.9	66.0 \pm 2.5	69.2 \pm 2.4

Table III provides a comparison of the three methods in terms of the kernel SVM-based classifier training times that are measured on a laptop with an Intel Core i5 (2.4 GHz) CPU and 4 GB RAM. In the Sparse-TDA method, the training times start from the computation of PIs. Not surprisingly, both of our method variants are usually much faster than kernel TDA as they use smaller sets of selected features. As expected, the reduction in training time is greater with linear weighting than with non-linear weighting. In fact, the Sparse-TDA method with linear weighting achieves about 65X speed-up for the SHREC'14 synthetic data set and roughly 8X speed-up for the OuTeX data set. However, there is no consistent speed-up for the SHREC'14 real data set owing to the fact that there are only 4-5 points in each PD in this case, rendering the training of the kernel TDA method exceptionally fast.

TABLE III
COMPARISON OF CLASSIFIER TRAINING TIMES (IN S) WITH THE SAME
TRAINING AND TEST DATA SET SPLIT AS IN THE ORIGINAL MULTI-SCALE
KERNEL TDA ARTICLE

Data Set	Sparse-TDA		Kernel TDA
	Linear weighting	Non-linear weighting	
SHREC'14 Synthetic	18.2 \pm 0.95	43.6 \pm 7.09	1220 \pm 27.9
SHREC'14 Real	55.4 \pm 1.1	160 \pm 54.4	107.8 \pm 5.4
OuTeX Texture	693 \pm 82.4	834 \pm 127	5457 \pm 979

Fig. 2 and Fig. 3 show the trends in improving the classification accuracy and reducing the classifier training time, respec-

¹<https://github.com/DIPHA/dipa>

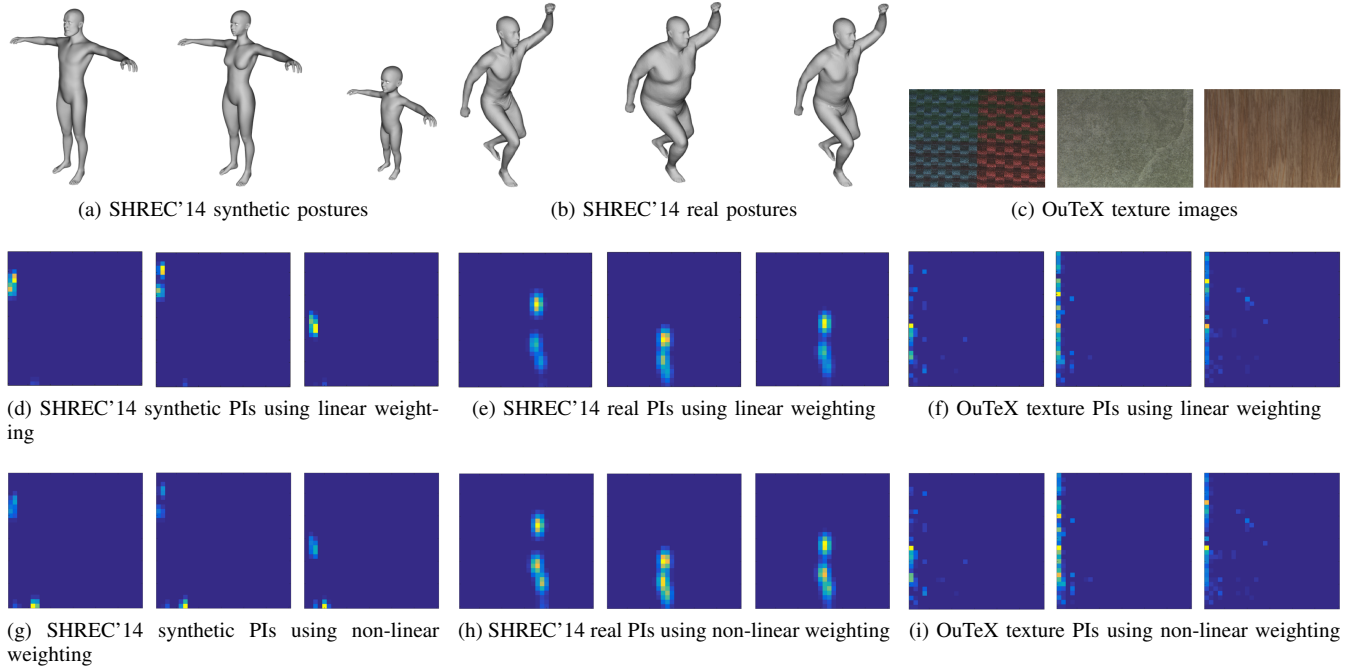


Fig. 1. Representative classes and corresponding persistence images (PIs) for various benchmark data sets using two Sparse-TDA methods.

tively, as a function of increasing training/testing split for all the benchmark data sets. Consistent with the results reported in Tables II, our classification accuracy is marginally inferior to that of the kernel TDA method for the SHREC'14 synthetic and OuTeX data sets. However, the Sparse-TDA method with linear and non-linear weighting marginally outperform the kernel method for the most challenging SHREC'14 real data set. The training time trends are also very similar to the results presented earlier in Table III, with more than an order of magnitude reduction for the SHREC'14 synthetic data set, comparable values for the SHREC'14 real set, and significant reduction for the OuTeX data set. The increase in classifier training times with higher training/test splits is, however, slightly more for both the variants of our method as compared to the kernel TDA method.

Overall, we observe that, for each of the benchmark data sets, at least one of the variants of our Sparse-TDA method outperforms the multi-scale kernel TDA method either in terms of classification accuracy or classifier training time.

VI. CONCLUSIONS

In this paper, we present a new method, referred to as the Sparse-TDA algorithm, that provides a sparse realization of a topological data analysis (TDA) algorithm. More specifically, we combine an optimal sparse sample selection method based on QR factorization using column pivoting with a state-of-the-art TDA method. Instead of persistence diagrams, we use a vector-based representation of persistent homology, called persistence images, with two different weighting functions to extract the topological features.

The results are promising on three benchmark multi-way classification problems pertaining to 3D meshes of human posture recognition, both for real and synthetic shapes, and

image texture detection. While our method gives comparable classification accuracy to the multi-scale kernel method, it results in a significant reduction of the classifier training times. This reduction is expected to lay the foundation for on-line adaptation of TDA on challenging data sets, which have a large number of classes, in response to changes in training sample availability.

In the future, we would like to further improve the accuracy of the Sparse-TDA method by designing our own weighting function for the persistence images. We would also like to come up with theoretical performance guarantees based on the characteristics of the data sets, particularly the training sample size for each individual class. Last but not the least, we plan to show the effectiveness of our method on other hard classification problems arising in robot visual perception and human face recognition.

ACKNOWLEDGMENTS

We would like to thank The Boeing Company for sponsoring this work in part under contract # SSOW-BRT-W0714-0004 and Dr. Tom Hogan for helpful discussions. The views and opinions expressed in the paper are, however, solely of the authors and do not necessarily reflect those of the sponsor.

REFERENCES

- [1] J. Reininghaus *et al.*, "A stable multi-scale kernel for topological machine learning," in *Proc. IEEE Conf. Comp. Vis. Pattern Recog. (CVPR 15)*, Boston, MA, June 2015, pp. 4741–4748.
- [2] M. Gao *et al.*, "Segmenting the papillary muscles and the trabeculae from high resolution cardiac CT through restoration of topological handles," in *Int. Conf. Info. Process. in Medical Imaging*. Springer, 2013, pp. 184–195.
- [3] M. Gameiro *et al.*, "A topological measurement of protein compressibility," *Japan J. Ind. and Appl. Math.*, vol. 32, no. 1, pp. 1–17, 2015.

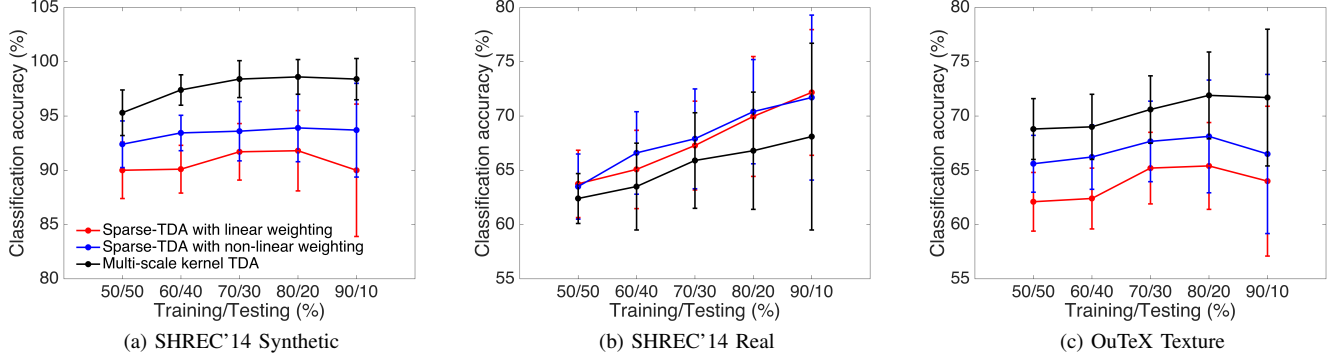


Fig. 2. Comparison of classification accuracy between sparse-TDA and state-of-the-art multi-scale kernel TDA methods for various training-testing data set partition ratios.

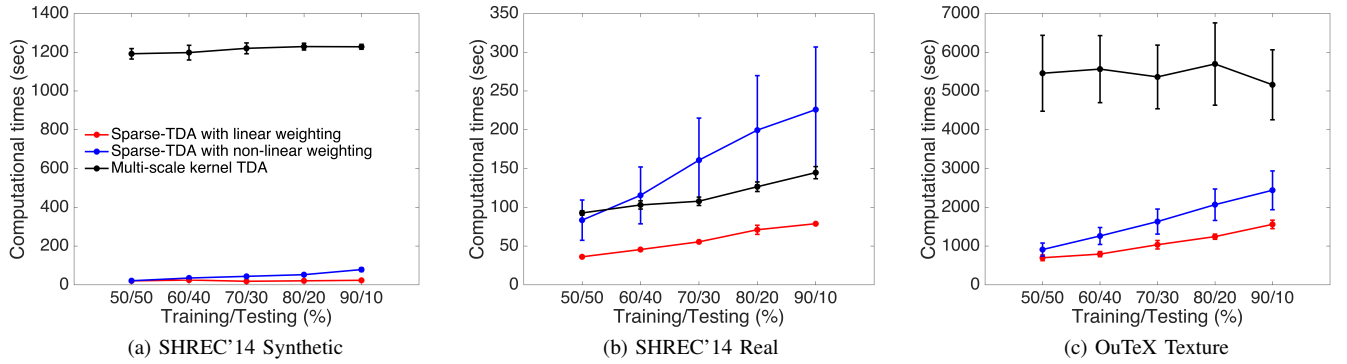


Fig. 3. Comparison of classifier training times between Sparse-TDA and state-of-the-art multi-scale kernel TDA methods for various training-testing data set partition ratios.

- [4] M. K. Chung *et al.*, "Persistence diagrams of cortical surface data," in *Int. Conf. on Info. Process. in Medical Imaging*, 2009, pp. 386–397.
- [5] Y. Hiraoka *et al.*, "Hierarchical structures of amorphous solids characterized by persistent homology," *Proc. Nat. Acad. of Sci.*, vol. 113, no. 26, pp. 7035–7040, 2016.
- [6] H. Edelsbrunner *et al.*, "Topological persistence and simplification," *Discrete & Computational Geometry*, vol. 28, no. 4, pp. 511–533, 2002.
- [7] A. Zomorodian and G. Carlsson, "Computing persistent homology," *Discrete & Computational Geometry*, vol. 33, no. 2, pp. 249–274, 2005.
- [8] D. Pachauri *et al.*, "Topology-based kernels with application to inference problems in Alzheimer's disease," *IEEE Trans. Med. Imag.*, vol. 30, no. 10, pp. 1760–1770, 2011.
- [9] P. Bubenik, "Statistical topological data analysis using persistence landscapes," *J. Mach. Learning Research*, vol. 16, no. 1, pp. 77–102, 2015.
- [10] G. Kusano *et al.*, "Persistence weighted gaussian kernel for topological data analysis," in *Proc. 33rd Int. Conf. Mach. Learning*, 2016.
- [11] H. Adams *et al.*, "Persistence images: A stable vector representation of persistent homology," *J. Mach. Learning Research*, to be published.
- [12] M. Carrière *et al.*, "Stable topological signatures for points on 3D shapes," in *Proc. Eurographics Symp. Geometry Process.*, vol. 34, no. 5, 2015, pp. 1–12.
- [13] T. Bonis *et al.*, "Persistence-based pooling for shape pose recognition," in *Int. Workshop on Comput. Topology in Image Context*, 2016, pp. 19–29.
- [14] M. Barrault *et al.*, "An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations," *Comptes Rendus Mathématique*, vol. 339, no. 9, pp. 667–672, 2004.
- [15] S. Chaturantabut and D. C. Sorensen, "Nonlinear model reduction via discrete empirical interpolation," *SIAM J. Sci. Comput.*, vol. 32, no. 5, pp. 2737–2764, 2010.
- [16] Z. Drmač and S. Gugercin, "A new selection operator for the discrete empirical interpolation method—improved a priori error bound and extensions," *SIAM J. Sci. Comput.*, vol. 38, pp. A631–A648, 2015.
- [17] S. Sargsyan *et al.*, "Nonlinear model reduction for dynamical systems using sparse sensor locations from learned libraries," *Physical Review E*, vol. 92, no. 033304, 2015.
- [18] K. Manohar *et al.*, "Environmental identification in flight using sparse approximation of wing strain," *Under revision for J. Fluids and Structures*, 2016.
- [19] B. W. Brunton *et al.*, "Sparse sensor placement optimization for classification," *SIAM J. on Appl. Math.*, vol. 76, no. 5, pp. 2099–2122, 2016.
- [20] D. Cohen-Steiner *et al.*, "Stability of persistence diagrams," *Discrete & Computational Geometry*, vol. 37, no. 1, pp. 103–120, 2007.
- [21] —, "Lipschitz functions have L_p -stable persistence," *Found. of Computational Math.*, vol. 10, no. 2, pp. 127–139, 2010.
- [22] T. Hofmann *et al.*, "Kernel methods in machine learning," *The Ann. of Stat.*, pp. 1171–1220, 2008.
- [23] D. Pickup *et al.*, "SHREC'14 track: Shape retrieval of non-rigid 3D human models," in *Proc. 7th Eurographics workshop on 3D Object Retrieval, EG 3DOR*, vol. 14, 2014.
- [24] T. Ojala *et al.*, "OuTeX - new framework for empirical evaluation of texture analysis algorithms," in *Proc. Int. Conf. Pattern Recog. (ICPR'02)*, Quebec City, QC, Canada, pp. 701–706.
- [25] J. Sun *et al.*, "A concise and provably informative multi-scale signature based on heat diffusion," in *Proc. Eurographics Symp. Geometry Process.*, vol. 28, no. 5, 2009, pp. 1383–1392.
- [26] Z. Guo *et al.*, "A completed modeling of local binary pattern operator for texture classification," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1657–1663, 2010.
- [27] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Trans. Intell. Sys. and Tech. (TIST)*, vol. 2, no. 27, 2011.